Lifetime Data Anal (2007) 13:449–462
DOI 10.1007/s10985-007-9047-z

Observational studies, clinical trials, and the women's health initiative

Ross L. Prentice

Received: 5 July 2007 / Accepted: 16 July 2007 / Published online: 18 October 2007

Abstract The complementary roles fulfilled by observational studies and randomized controlled trials in the population science research agenda is illustrated using results from the Women's Health Initiative (WHI). Comparative and joint analyses of clinical trial and observational study data can enhance observational study design and analysis choices, and can augment randomized trial implications. These concepts are described in the context of findings from the WHI randomized trials of postmenopausal hormone therapy and of a low-fat dietary pattern, especially in relation to coronary heart disease, stroke, and breast cancer. The role of biomarkers of exposure and outcome, including high-dimensional genomic and proteomic biomarkers, in the elucidation of disease associations, will also be discussed in these same contexts.

Keywords Biomarker · Cohort study · Diet · Genomics · Hormones · Proteomics · Randomized controlled trial

1 Introduction

It is a great pleasure to provide a contribution to this issue in honor of my esteemed colleague, Dr. Norman Breslow. Norm was one of the few Seattleites I knew before my family moved to Seattle in 1974. He was already a key member of the Biostatistics Department at the University of Washington at that time, a department formed in 1970 and capably led by Drs. Ed Perrin, Donovan Thompson, Norm Breslow, Tom Fleming, and Bruce Weir over the years

R. L. Prentice (✉)
Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA
e-mail: rprentic@fhcrc.org

since its inception. I moved to Seattle to be an initial on-site biostatistician at the then-fledgling Fred Hutchinson Cancer Research Center, having developed an interest in biomedical applications while visiting Drs. Marvin Zelen and Jack Kalbfleich, then at the State University of New York at Buffalo. I asked Norm about the potential of the start-up Hutchinson Center. He responded that the research groups headed by Drs. E. Donnall Thomas, Karl Erik, and Ingegard Hellstrom, and Robert Nowinski that were to be based at the Hutchinson Center were 'absolutely tops.' I interviewed for the position and soon was en route to Seattle. Though based at the Hutchinson Center, the University of Washington was my principal academic home during my initial years in Seattle. Norm and I had common interests in observational study methods and clinical trials. We wrote a few papers together, and I learned a lot by observing Norm's thorough approach to research and his ability to explain technically difficult concepts clearly to diverse audiences. Such overlapping interests continue to this day, and it has always been important to me to know that Norm and other valued colleagues have both appreciation for, and many valuable contributions to, observational study and clinical trial methodology and application.

Norm's two-volume series with Dr. Nick Day on case-control and cohort study methods are classics among epidemiologists and biostatisticians (Breslow and Day 1980, 1987). Norm's several decades of involvement and leadership in the National Wilms Tumor Study, in collaboration with Drs. Giulio D'Angio, Audrey Evans, and others, serves as a prime example of randomized controlled trial and observational study integration, and of impressive progress in the management of this important childhood tumor.

2 Scope of this contribution

The Women's Health Initiative (WHI), in which I have been engaged since its inception in 1992, also integrates an observational study (OS) and a multifaceted randomized controlled trial (CT). The WHI is conducted among postmenopausal women, in the age range 50–79 at the time of enrollment during 1993–1998, seen at one of 40 clinical centers throughout the United States. The CT involved four distinct randomized controlled intervention evaluations in a partial factorial design (WHI Study Group 1998). Two of these involved postmenopausal hormone therapy (HT) versus placebo comparisons, either conjugated equine estrogen alone among women who were post-hysterectomy, or this same estrogen preparation plus medroxyprogesterone acetate among women with a uterus. These trials projected a major reduction in coronary heart disease risk, with a breast cancer risk elevation also anticipated. To enroll in the CT women could enroll in one of the HT trials or the dietary modification (DM) trial, or both. The DM trial studied a low-fat dietary pattern for the prevention of cancer, with breast and colorectal cancer separately as the designated primary outcomes. At the one-year anniversary from initial randomization, eligible women in the CT were invited to be further randomized into a calcium and vitamin D versus placebo supplementation trial, primarily aimed to reduce

the risk of fracture, with hip fracture the designated primary outcome. A total of 68,132 women enrolled in the CT. This total is 60.6% of the total sample sizes for the four CT components, providing a cost and logistics justification for the use of a partial factorial design with overlapping components.

The WHI also includes a cohort study among 93,676 postmenopausal women in the age range 50–79 at enrollment. This OS enrolled women from the same populations, over essentially the same time period, with major elements of clinical outcome ascertainment and much risk factor and exposure data collection common to the two cohorts. The common data collection included an in-person interview to assess the use and timing of exogenous hormones prior to WHI enrollment.

This contribution reviews our attempts to date to explain apparent discrepancies between randomized controlled trial results from the CT, with observational associations from WHI cohorts, both for the HT trials (Sect. 3), and the DM trial (Sect. 4). Data collection, including that for high-dimensional genomic and proteomic data, to help understand health effects observed in the CT are also briefly described (Sect. 5). The presentation ends with some brief comments on the future preventive intervention research agenda, and on related methodology needs (Sect. 6).

3 Women's health initiative studies of postmenopausal hormone therapy

3.1 Estrogen plus progestin trial and cardiovascular disease

As mentioned above, two of the CT components evaluated postmenopausal hormone therapy, a continuous regimen of 0.625 mg per day of conjugated equine estrogens (CEE) among 10,739 women who were post-hysterectomy at enrollment (E-alone trial), and a continuous regimen of the same estrogen plus 2.5 mg per day of medroxyprogesterone acetate (MPA) among 16,608 women who were with uterus at enrollment (E+P trial). These preparations were used by about 8 million and 6 million women, respectively in the United States, and many more millions of women worldwide until the E+P trial was stopped early in 2002 (Writing Group for the WHI Investigators 2002), when it was judged that risks exceeded benefits over an average 5.6-year intervention period. The left side of Table 1 shows hazard ratio estimates (HRs) and nominal 95% confidence intervals at the time of early stopping. These HRs were based on Cox (1972) regression analysis, with baseline hazard ratio stratified on age (5-year age group), prior history of the disease under evaluation, and randomization in the dietary modification (DM) trial component, to be described below. The fact that there was a moderate increase in the designated primary outcome, coronary heart disease (CHD), while a major reduction had been hypothesized was a cause for much reaction by gynecologists, cardiovascular epidemiologists, and other clinical groups. E+P trial results also led to discussion of the details of the WHI study (e.g., characteristics of enrollees; medication adherence; monitoring and reporting methods), and of the reliability of a very extensive observational

Table 1 Clinical outcomes in the WHI postmenopausal hormone therapy trials[a]

Outcomes	E+P trial Hazard ratio	95% CI	E-alone trial Hazard ratio	95% CI
Coronary heart disease	1.29	1.02–1.63	0.91	0.75–1.12
Stroke	1.41	1.07–1.85	1.39	1.10–1.77
Venous thromboembolism	2.11	1.58–2.82	1.33	0.99–1.79
Invasive breast cancer	1.26	1.00–1.59	0.77	0.59–1.01
Colorectal cancer	0.63	0.43–0.92	1.08	0.75–1.55
Endometrial cancer	0.83	0.47–1.47		
Hip fracture	0.66	0.45–0.98	0.61	0.41–0.91
Death due to other causes	0.92	0.74–1.14	1.08	0.88–1.32
Global index	1.15	1.03–1.28	1.01	0.91–1.12
Number of women	8,506	8,102	5,310	5,429
Follow-up time, mean (SD), mo	62.2 (16.1)	61.2 (15.0)	81.6 (19.3)	81.9 (19.7)

[a]Writing Group for the WHI Investigators (2002), WHI Steering Committee (2004)

literature (e.g., Stampfer and Colditz 1981; Barrett-Connor and Grady 1998) that mostly suggested a CHD risk that is 40–50% lower among E-alone or E+P users compared to non-users.

In an attempt to understand the basis for results that appeared to be discrepant between the WHI E+P trial and related observational research, a group of WHI investigators (Prentice et al. 2005b) contrasted results from the CT and OS. For this purpose a subset of 53,054 women from the OS were selected who were with uterus and not using unopposed estrogens (i.e., estrogen without progestin) at the time of WHI enrollment. The ratio of age-adjusted CHD incidence rates for (baseline) E+P users compared to non-users of hormone therapy was less than one in the OS, and only about 50–60% of the corresponding HR from the E+P trial. Similarly, the E+P HRs for stroke and venous thromboembolism in the OS were only 50–60% of those from the E+P trial. After control for an extensive set of potential confounding factors, the HRs in the OS remained about 30–40% lower ($p < 0.05$) than those in the E+P trial for each of the three diseases. However, upon control for time from E+P initiation, defined as time from enrollment in the CT and in the non-user group in the OS, and as time from enrollment plus the duration of the baseline E+P episode at enrollment for the E+P user group in the OS, the hazard ratios from the two cohorts were in considerably better agreement, with differences that could be explained readily by chance. For example, upon control for time from E+P initiation as a time-dependent variable, and for potential confounding factors in stratified Cox models, the ratio of E+P HR in the OS to E+P HR in the CT (95% confidence interval (CI)) as a multiplicative HR factor was estimated as 0.93 (0.64, 1.36). Hence, the residual 7% lower estimated HR in the OS compared to the CT is readily explained by chance. The corresponding OS to CT ratio of HRs (95% CI) was 0.76 (0.49, 1.18) for stroke, and 0.84 (0.54, 1.28), leaving the possibility of some small residual bias, especially for stroke.

The principal source of discrepancy between the WHI CT and OS in regard to E+P and cardiovascular disease appeared to derive from HRs that were

elevated early and declined after the first few years of use. For CHD, the estimated HRs from combined CT and OS data analysis were 1.56, 1.16, and 0.81 in the first 2 years, 2–5 years, and more than 5 years from E+P initiation categories, respectively. The WHI OS, like many other cohort studies, enrolled women who may have already used E+P for some years, so that most HR information pertained to the more than 5 year from E+P initiation category. The CT, on the other hand, provides an assessment of HRs from the time of E+P initiation, and with an average follow-up of 5.6 years most HR information pertains to the first few years of use. The CT and OS hence each provide valuable, substantially complementary, information. Contrasting CT and OS findings identified an important potential source of bias in observational studies, and has implications for study design and analysis. Specifically, an observational study of an exposure such as E+P needs to be sure to include a sufficient number of recent E+P initiators to be in a position to assess risks and benefits from the beginning of exposure. Similarly, a proportional hazards analysis assumption, while unlikely to give seriously misleading results in the randomized trial setting, can have a major impact in observational studies if there is substantial nonproportionality and 'late entry' into the cohort, as in the WHI OS. Proportional hazards or related assumptions should be tested to the extent possible using available data, and relaxed as necessary. Also, joint analyses of clinical trial and observational study findings may be able to add precision and augment clinical trial findings, if the two sources can be brought into agreement. For example, the time-varying HRs for E+P and CHD listed above can be used to give an estimated average HR over a 10-year period from E+P initiation (95% CI) of 1.07 (0.92, 1.24), suggesting the absence of important benefit even over a fairly lengthy usage period. See Prentice et al. (2005a) for additional related discussion, and for comments by several leading biostatisticians and epidemiologists.

3.2 Estrogen-alone trial and cardiovascular disease

The WHI E-alone trial also stopped early in 2004 (WHI Steering Committee 2004), primarily based on a stroke elevation of a similar magnitude to that for E+P. As shown on the right side of Table 1, the risks and benefits of this CEE regimen were fairly well-balanced over an average 7.1-year follow-up period, with a 'global index' defined as time to the earliest of the outcomes listed above it, having a HR (95% CI) of 1.01 (0.91, 1.12). The HR (95% CI) for the primary CHD outcome was 0.91 (0.75, 1.12) at the time of stopping, so that the anticipated major benefit did not arise, not was it likely to do so by the planned termination about 1 year later.

WHI investigators (Prentice et al. 2006) undertook similar HR comparisons between the E-alone trial and the subset of 38,313 women from the OS who were post-hysterectomy and not taking E+P preparations at the time of WHI enrollment. Once again, age-adjusted HRs from the OS were only 50–60% of those for the CT, but agreed fairly closely after controlling for confounding factors and, especially, for time from E-alone initiation. A novel aspect of these

analyses used the comparative HRs for E+P in the CT and OS to formulate a residual bias estimator in analysis of E-alone in the OS. The resulting E-alone HRs from these three cohorts (E-alone OS, E+P OS, E+P CT) agreed fairly well with E-alone HRs from the CT, as was also the case when the roles of E-alone and E+P were reversed.

3.3 Hormone therapy trials and breast cancer

The trigger for the early stoppage of the E+P trial was an elevation in the 'primary safety' outcome of invasive breast cancer. WHI investigators have carried out similar contrasts of HRs from the CT and OS for invasive breast cancer for both E+P and E-alone. The breast cancer HRs shown in Table 1 are considerably lower than those from an extensive observational literature. For example, the UK Million Women's Study estimated HRs (95% CIs) of 2.00 (1.88, 2.12) for E+P regimens and 1.30 (1.20, 1.40) for E-alone regimens, though HRs are somewhat lower than these in other observational studies. The WHI CT and OS breast cancer analyses have yet to be published, and so won't be discussed in any detail here. These, however, yield larger HRs in the OS versus the CT for both E+P and E-alone, differences that are not explained by available confounding factors, or time from hormone therapy initiation. Efforts to understand the important residual bias led to the identification of an important effect modifying factor that is helping to clarify both methodologic aspects and clinical implications for these preparations, thereby providing an additional example of the valuable interplay between trials and observational studies.

4 Women's health initiative studies of a low-fat dietary pattern

4.1 Trial design and breast cancer findings

Another important component of the WHI CT tests the hypothesis that a low-fat dietary pattern can reduce cancer incidence, with invasive breast cancer, and invasive colorectal cancer as designated primary outcomes. The Dietary Modification (DM) trial randomized 48,835 postmenopausal women either to a low-fat eating pattern (40%), or to usual diet (60%). The intervention group women were taught by nutritionists, in small groups of size 8–15, how to make and maintain a change to a low-fat dietary pattern having goals of not more than 20% of energy (calories) from fat, five or more servings of fruit and vegetables per day, and six or more grain servings per day. The intervention included nutritional strategies to identify the fat content in food, to limit fat in food preparation, and to budget fat consumption, as well as behavioral strategies related to self-management and self-efficacy, social support, and relapse prevention.

The low-fat diet and breast cancer hypothesis dates back to animal feeding experiments in the 1940s, which were followed by supportive international

correlative and national time trend studies (e.g., Prentice and Sheppard 1990). The hypothesis was also supported by a combined analysis of 12 case–control studies (e.g., Howe et al. 1990), that used a variety of dietary assessment methods in diverse populations, but not by an analysis of seven cohort studies (Hunter et al. 1996), that took place in Western populations and each used a food frequency questionnaire (FFQ). The hypothesis became rather controversial during the 1980s as these cohort data began to emerge, with some arguing that an expensive randomized trial, as in the WHI, was not justified. At the heart of this argument is the issue of the measurement properties of dietary assessment methodologies, and the corresponding reliability of observational studies in nutritional epidemiology. Similar issues are pertinent also to physical activity epidemiology and to energy balance and obesity studies.

It is clear from repeat application of available dietary assessment tools, such as FFQs, food records or dietary recalls, that there is a substantial random measurement error in the assessment of nutritional factors, such as total fat consumption or percent of energy from fat. What is less clear, and has not been acknowledged in nutritional epidemiology studies to date, is the importance of a systematic component to consumption measurement errors. For example, Heitmann and Lissner (1995) show that obese women and men substantially underestimate total energy consumption on self-report, whereas such underestimation is minimal among slim persons. Moreover, the systematic bias associated with body mass was greater for total energy than for protein energy. These studies reflect the availability of good biomarkers of total energy expenditure (a doubly labeled water measure) and protein expenditure (urinary nitrogen-based measure). Such biomarkers have not yet been established for other macronutrients, or for most micronutrients. A systematic component to measurement error could relate to many study subject characteristics and behaviors, and could substantially distort dose-response relationships in observational studies. Hence, the lack of acknowledgement of systematic measurement error casts a major shadow over the reliability of the existing body of nutritional epidemiology literature, and provides incentive for randomized controlled trials of nutritional hypotheses having substantial public health potential.

Based primarily on international correlation and migrant data, the DM trial design hypothesized a HR for a 20% versus a 40% energy from fat diet that reduced linearly from one at randomization to a minimum of 0.5 at 10 years following randomization. An effect of this magnitude would have great public health potential as it would imply, for example, a 37.5% reduction in breast cancer risk after a decade of a population change from a 35% energy from fat diet to a 20% energy from fat diet. However, the lengthy period of time (10 years) required for a full intervention effect under this hypothesis, and the fact that the intervention and comparison groups were hypothesized to differ in percent energy from fat by only 13% at 1 year and 11% at 10 years, along with some provision for loss to follow-up and deaths due to competing risks, led to a projected breast cancer incidence in the intervention group that is only 14% lower than in the comparison group in the DM trial (WHI Study Group

1998). This type of calculation illustrates the challenge in assessing intervention effect when the HR is likely to decrease from unity slowly over an extended time period, and it illustrates the vulnerability of study power (projected to be 86% for breast cancer) to modest departures from design assumptions in the time course of the intervention effect, or in the realized adherence to dietary assumptions in intervention and comparison groups. To enhance the dietary difference between the intervention and comparison (usual diet) groups the DM trial used a FFQ to screen out approximately one-half of otherwise eligible women based on an estimated % energy from fat less than 32. This was expected to achieve a baseline % energy from fat in the vicinity of 38, as in preceding feasibility studies. As it turned out, however, this baseline % energy from fat, as assessed by the comparison group FFQs at 1 year from randomization, was only about 35%, and as a result, the % energy from fat difference between the randomization groups was only an estimated 10.7%, rather than projected 13%, at 1 year from randomization, decreasing to 8.1% at 6 years. The resulting projected breast cancer difference between the intervention and comparison groups under other design assumptions was only about 8–9%, as a result of a % energy from fat difference that was overall only about 70% of that projected and an 8.1-year rather than a 9-year average follow-up period. The principal DM trials results were published in 2006. For breast cancer (Prentice et al. 2006) the intervention group showed a breast cancer incidence that was 9% lower in the intervention group versus the comparison group with a HR (95% CI) of 0.91 (0.83, 1.01), which was suggestive of benefit, but not significant at conventional levels. Table 2 shows this and other key findings from the trial. The biological plausibility for a breast cancer risk reduction was enhanced by reductions in plasma estradiol concentrations that were about 15% larger in the intervention versus the comparison group. Intervention group women also reported a vegetable and fruit consumption that exceeded that in the comparison group by a little over one serving per day, and this was supported by corresponding plasma micronutrient differences. The strongest evidence for a breast cancer reduction in the intervention group arose from a significant ($P = 0.04$) interaction between baseline % energy from fat, as measured by 4-day food records, and the breast cancer hazard ratio, with women in the upper quartile of % energy from fat making larger % energy from fat reductions in the intervention group and having a HR (95% CI) of 0.78 (0.64, 0.95). A lower incidence in the intervention group was evident for difficult-to-manage progesterone receptor negative tumors, and particularly for estrogen receptor positive/progesterone receptor negative tumors.

The results just described were widely reported in the media as showing no effect of a low-fat diet on breast cancer. However, these findings are entirely consistent with the hypothesized intervention effect with its important public health potential, even though not precise enough to be certain of an intervention effect. WHI investigators and other colleagues are engaged in a number of additional studies in an attempt to further clarify the health-related potential of a low-fat diet, and to advance the nutritional and physical activity epidemiology research areas more generally.

Table 2 Risk of breast cancer and other major clinical outcomes by randomization assignment in the dietary modification trial

	Intervention group # of cases (Annualized %)	Comparison group # of cases (Annualized %)	Hazard ratio (95% CI)	Unweighted *P*-value	Weighted *P*-value
Breast cancer incidence	655 (0.42%)	1,072 (0.45%)	0.91 (0.83, 1.01)	0.07	0.09
Breast cancer mortality	27 (0.02%)	53 (0.02%)	0.77 (0.48, 1.22)	0.26	0.27
Total cancer incidence	1,946 (1.23%)	3,040 (1.28%)	0.96 (0.91, 1.02)	0.15	0.10
Total cancer mortality	436 (0.28%)	690 (0.29%)	0.95 (0.84, 1.07)	0.41	0.22
Total mortality	950 (0.60%)	1,454 (0.61%)	0.98 (0.91, 1.07)	0.70	0.29
Global index	2,051 (1.30%)	3,207 (1.35%)	0.96 (0.91, 1.02)	0.16	0.16

Prentice et al. (2006)

4.2 Dietary and Physical Activity Epidemiology Methods

For example, Freedman et al. (2006) compared the results of observational analyses of the fat and breast cancer association using FFQs and food records among the 29,294 women randomized to the DM comparison group. As in an earlier study in a UK cohort (Bingham et al. 2003) a positive association was observed between energy-adjusted fat consumption and breast cancer risk when diet was assessed using food records, but not when using FFQs. This reinforces the likelihood that the measurement properties of dietary assessment tools may have a crucial influence in nutritional epidemiology findings.

Biomarkers of nutrient consumption provide an important pathway to enhancing the reliability of observational studies in the nutritional epidemiology areas. For certain nutritional factors, including total energy, protein and sodium, the expenditure of the 'nutrient' over a short time period can be estimated from urinary recovery. Among weight stable persons these urinary measures provide intake assessments that are objective and plausibly adhere to a classical measurement model

$$\mathrm{X} = \mathrm{Z} + \mathrm{e}$$

where X is the biomarker, Z is the targeted consumption, and e is random noise that is independent of Z and of other study subject characteristics (e.g., body mass, age, ...). The corresponding self-report W from the same study subject can be allowed to have a more flexible measurement model (e.g., Prentice 1996; Carroll et al. 1998; Prentice et al. 2002)

$$\mathrm{W} = \mathrm{a}_0 + \mathrm{a}_1\mathrm{Z} + \mathrm{a}_2\mathrm{V} + \mathrm{a}_3\mathrm{Z} \otimes \mathrm{V} + \mathrm{r} + \varepsilon$$

where V is a vector of study subject characteristics that may relate to the systematic bias of W, r is a random effect that allows the measurement errors for replicates of W to be correlated, ε is an independent noise component, and $\{a_i\}$ are coefficients to be estimated. Observations $\{X, W\}$ on a subset of a study cohort including some replicates, leads via simple estimating equations to a calibration equation that allows each self-report value W to be replaced by an estimate $\hat{Z}$ of the underlying nutritional factor that has been corrected for measurement error. These 'calibrated' estimates can then be used in logistic or Cox regression to estimate odds ratios (Sugar et al. 2007) or hazard ratios (Shaw 2006) for the nutritional factor using regression calibration or other statistical procedures.

Biomarkers, including the doubly labeled water assessment of total energy consumption, urinary assessment of protein consumption, and various blood nutrient concentrations, have recently been collected for 544 women in the DM trial (50% intervention, 50% comparison), along with corresponding FFQs. Analyses of these data have yet to be published, but show strong evidence of systematic bias for total energy, with much more modest systematic bias for protein. These data are currently being used in analyses to examine the association between total energy and protein, and % energy from protein with the risk of various cancers, vascular diseases, and diabetes in the WHI cohorts. A second biomarker study has recently been initiated among 450 women in the WHI OS. This study has somewhat broader goals of calibrating frequencies, records and recalls of both diet and physical activity and applying the calibrated estimate in corresponding association studies.

5 Ongoing biomarker studies to explain intervention effects in the CT

Various types of biomarker studies have been, or are being, carried out toward elucidating the biological mechanisms underlying the postmenopausal hormone therapy and low-fat diet intervention effects noted above, as well as the effects of an additional CT intervention that involved calcium and vitamin D supplementation for fracture and colorectal cancer prevention (Jackson et al. 2006; Wactawski-Wende et al. 2006). These are mostly in the form of nested case-control studies that examine baseline biomarkers that may relate to the magnitude of intervention effects, or that examine biomarker changes following the initiation of intervention that may provide an explanation for some or all of observed intervention effects. For example, a cardiovascular disease case–control study has been completed within the E+P and E-alone trials. This study focused on markers of inflammation, coagulation, thrombosis and lipids, and included a number of genetic polymorphisms related to these processes.

Although these markers tend to relate to disease incidence in the anticipated fashion in this study, analyses reported to date of biomarker changes have not revealed much insight into such key intervention effects as the observed stroke and venous thromboembolism elevation with E+P and E-alone, or the early elevation in CHD with E+P.

A genome-wide association study of CHD, stroke, and breast cancer has been carried out in the OS and E+P trial cohorts with goals of identifying novel aspects of genotype as risk factors for these diseases, as well as of identifying interactions of genotype with E+P effects. This three-stage study (Prentice and Qi 2006) was conducted in collaboration with Drs. David Cox and Dennis Ballinger of Perlegen Sciences, using Perlegen's 360,000 tag-SNP set. The first two stages were conducted in the OS, the first involving 1,000 cases and 1,000 controls for each disease with matched case-control pools of size 125, and the second individual genotyping of about 10,000 SNPs meeting first stage statistical criteria (9,000) or included from candidate gene considerations (1,000) for each disease. The second stage involved 650–800 cases and controls, while the third stage involved individual genotyping of about 300 SNPs for each disease meeting earlier stage criteria among cases and matched controls in the E+P trial. The data from this study have only recently been assembled, and related publications yet to be prepared.

Toward assessing a fairly comprehensive set of proteomic changes in relation to hormone therapy WHI investigators are currently comparing the plasma proteomes between baseline and one year from randomization among 50 women adherent to E+P and 50 women adherent to E-alone in collaboration with Dr. Sam Hanash (Fred Hutchinson Cancer Research Center). Dr. Hanash's Intact Protein Analysis System (Wang et al. 2005) is capable of a quantitative contrast of plasma concentrations for about 1,000 individual proteins. Comparisons of baseline to 1-year plasma pools formed from 10 E+P or E-alone users are nearing completion at the time of this writing. Complementary proteomic analyses of plasma specimens from CHD, stroke, and breast cancer cases and matched controls are also underway. The concept is to choose a small number of novel candidate proteins from these two sources for specialized test development and application to cases and controls in the hormone therapy trials in an attempt to more fully understand hormone therapy effects on these diseases.

The central issues relative to the DM trial are more closely related to assessing, rather than explaining, intervention effects. Nevertheless, the nutritional biomarker studies previously mentioned are already contributing to the understanding of intervention effects in relation to dietary data reported on the FFQ. For example, these data help to align the measured weight change data with intervention versus comparison group differences in FFQ energy consumption estimates. Also, the completed, and especially the ongoing, biomarker studies include indirect calorimetry to assess resting energy expenditure which, in conjunction with doubly labeled water assessments of total energy expenditure, lead to an objective assessment of activity-related energy expenditure for use in calibrating self-reported physical activity data. As a byproduct, indirect calorimetry also yields a respiratory quotient which reflects the fat, carbohydrate, and protein composition of recent diet. These data, in conjunction with total energy and protein energy biomarkers, will be used to develop biomarker estimates of fat and carbohydrate consumption for use in calibrating fat and carbohydrate self-reports in nutrient association studies, and in analyses

to examine the extent to which changes in the fat composition of the diet can explain breast cancer or other effects of the DM intervention.

6 The future research agenda, and concluding remarks

Chronic disease population research is a vast and important research arena with many outstanding methodologic topics. Primary prevention trials provide a crucial research tool in this agenda but, unlike therapeutic research, the cost and logistical complexities are such that only a few full-scale primary prevention trials can be conducted at any point in time. This implies a continuing important role for observational studies, and the need to develop the methods to rigorously assess the reliability of cohort and case–control studies, and to enhance study design, conduct, and analysis, as necessary. Settings, such as the WHI where both types of studies are available, illustrate the potential for important bias in observational studies as currently conducted, and provide insights as to improvements to strengthen both study types.

In diet and physical activity epidemiology there is mounting evidence that measurement error in exposure assessment may greatly influence study findings. Measurement modeling and accommodation has some strong statistical components, and statisticians need to be involved, not just in data analysis, but also in helping to formulate the measurements that are needed and related study designs. This is especially the case for nutrients, such as many micronutrients, where a recovery biomarker that plausibly adheres to a classical measurement model has not yet been developed, and only plasma nutrient concentrations are available. Data external to the study cohort, such as data from pertinent human feeding studies are evidently required, under these circumstances, to allow such biomarkers to be used to calibrate corresponding self-report nutrient consumption estimates, but necessary methods have yet to be developed.

The rather large research area of biomarkers to explain intervention effects or exposure-disease associations is likewise in need of focused methodologic development. Methodologic needs include issues of using high-dimensional biomarkers and related study designs, which are receiving considerable statistical attention, and also fundamental issues of assessing when and if a biomarker, or group of biomarkers, provides an adequate explanation for an important observed intervention effect. For example, an assessment is needed of the reliability of methods that examine the impact of the addition of a biomarker change in a regression model, on the magnitude to an intervention effect, with emphasis on such factors as the nature of the intervention effect on the biomarker, and biomarker measurement error properties.

Finally, there is much need for an enhanced preventive intervention development enterprise. While good intervention concepts may arise from observational epidemiology and from therapeutic trials, these sources leave a very large gap and there is a need for specialized and vigorous intervention developments using biomarker outcomes. For example, desirable dietary and physical activity patterns may be studied in small-scale human feeding and exercise trials in

relation to proteomic and other biomarker changes. As the knowledge base develops to relate such biomarker changes to the risk of a spectrum of chronic diseases, one may be in a position to identify practical interventions having a favorable benefit versus risk profile, for consideration as interventions for study in full-scale prevention trials. This type of infrastructure may also serve to avoid the late discovery of adverse effects when products having preventive potential are eventually tested in full-scale trials.

While these types of methodologic challenges may seem daunting, it is instructive to consider the extent of epidemiologic and clinical trial methods developments over the roughly four decades of Norm Breslow's career. While Mantel–Haenszel methods and Kaplan–Meier curves were available in the late 1960s, Cox regression methods, case-control and within-cohort sampling techniques and related regression methods, covariate measurement error methods, and genetic epidemiology methods, were mostly yet to be conceived. We look forward to comparable developments and progress in upcoming years by Norm and other valued colleagues worldwide, to advance the public health and population science research agenda.

Acknowledgements This work was supported by HIH grants CA53996, CA86368, and CA106320, and by contract WH-2-2110 from the National Heart, Lung, and Blood Institute.

References

Barrett-Conner E, Grady D (1998) Hormone replacement therapy, heart disease, and other considerations. Annu Rev Public Health 19:55–72

Bingham SA, Luben R, Welch A, Wareham N, Khaw KT, Day N (2003) Are imprecise methods obscuring a relationship between fat and breast cancer?. Lancet 362:212–214

Breslow NE, Day NE (1980) Statistical methods in cancer research—the analysis of case-control studies, vol 1. IARC Scientific Publications 32, International Agency for Research on Cancer, Lyon, France

Breslow NE, Day NE (1987) Statistical methods in cancer research—the design and analysis of cohort studies, vol 2. IARC Scientific Publications 32, International Agency for Research on Cancer, Lyon, France

Carroll RJ, Freedman L, Kipnis V, Li L (1998) A new class of measurement error models, with application to dietary data. Can J Stat 26:467–477

Cox DR (1972) Regression models and life tables (with discussion). J Royal Statistical Soc B 34:187–220

Freedman LS, Potischman N, Kipnis V, Midthune D, Schatzkin A, Thompson FE, Troiano RP, Prentice R, Patterson R, Carroll R, Subar AF (2006) A comparison of two dietary instruments for evaluating the fat-breast cancer relationship. Int. J. Epidemiol. 35(4):1011–1021 [Epub May 3, 2006]

Heitmann BL, Lissner L (1995) Dietary underreporting by obese individuals—is it specific or non-specific? Br Med J 311:986–989

Howe GR, Hirohata T, Hislop TG, Iscovich JM, Yuan JM, Katsouyanni K, Lubin F, Marubini E, Modan B, Rohan T et al (1990) Dietary factors and risk of breast cancer: combined analysis of 12 case–control studies. J Nat Cancer Inst 82:561–569

Hunter D, Spiegelman D, Adami H-O, Beeson L, van den Brandt PA, Folsom AR, Fraser GE, Goldbohm RA, Graham S, Howe GR et al (1996) Cohort studies of fat intake and the risk of breast cancer—a pooled analysis. New England J Med 334:356–361

Jackson RD, LaCroix AZ, Gass M, Wallace RB, Robbins J, Lewis CE, Bassford T, Beresford SA, Black HR, Blanchette P et al (2006) The Women's Health Initiative trial of calcium plus vitamin D supplementation on risk for fractures. N Engl J Med 354:669–683

Prentice RL (1996) Measurement error and results from analytic epidemiology: dietary fat and breast cancer. J Nat Cancer Inst 88(23):1738–1747

Prentice RL, Caan B, Chlebowski R, Patterson R, Kuller LH, Ockene JK, Margolis KL, Limacher MC, Manson JE, Parker LM et al (2006) Low-fat dietary pattern and risk of invasive breast cancer: the Women's Health Initiative randomized controlled dietary modification trial. J Am Med Assoc 295:629–642

Prentice RL, Pettinger M, Anderson GL (2005a) Statistical issues arising in the Women's Health Initiative (with discussion). Biometrics 61:899–941

Prentice RL, Langer R, Stefanick ML, Howard BV, Pettinger M, Anderson G, Barad D, Curb JD, Kotchen J, Kuller L, Limacher M, Wactawski-Wende J (2005b) Combined postmenopausal hormone therapy and cardiovascular disease: toward resolving the discrepancy between observational studies and the Women's Health Initiative clinical trial. Am J Epidemiol 162:1–11

Prentice RL, Langer RD, Stefanick ML, Howard BV, Pettinger M, Anderson GL, Barad D, Curb JD, Kotchen J, Kuller L, Limacher M, Wactawski-Wende J (2006) Combined analysis of Women's Health Initiative observational and clinical trial data on postmenopausal hormone treatment and cardiovascular disease. Am J Epidemiol 163(7):589–599

Prentice RL, Qi L (2006) Aspects of the design and analysis of high-dimensional SNP studies for disease risk estimation. Biostatistics 7:339–354

Prentice RL, Sheppard L (1990) Feasibility and importance of clinical trials of a low fat diet to reduce cancer risk. Prog Clin Biol Res 346:205–215

Prentice RL, Sugar E, Wang CY, Neuhouser M, Patterson R (2002) Research strategies and the use of nutrient biomarkers in studies of diet and chronic disease. Public Health Nutr 5:977–984

Shaw P (2006) Estimation methods for Cox regression with nonclassical measurement error. Doctoral Dissertation Department of Biostatistics, University of Washington

Stampfer M, Colditz G (1991) Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. Prev Med 20:47–63

Sugar EA, Wang CY, Prentice RL (2007) Logistic regression with exposure biomarkers and flexible measurement error. Biometrics 63:143–151

Wactawski-Wende J, Kotchen JM, Anderson GL, Assaf AR, Brunner RL, M. J. O'Sullivan, Margolis KL, Ockene JK, Phillips L, Pottern L et al (2006) Calcium, vitamin D supplementation and colorectal cancer in postmenopausal women: the Women's Health Initiative clinical trial. N Eng J Med 354:684–696

Wang H, Clouthier SG, Galchev V, Misek DE, Duffner U, Min CK, Zhao R, Tra J, Omenn GS, Ferrara JL, Hanash SM (2005) Intact-protein-based high-resolution three-dimensional quantitative analysis system for proteome profiling of biological fluids. Mol and Cell Proteomics 4:618–625

Women's Health Initiative Steering Committee (2004) Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. J Am Med Assoc 291(14):1701–1712

Women's Health Initiative Study Group (1998) Design of the Women's Health Initiative clinical trial and observational study. Control Clin Trials 19:61–109

Writing Group for the Women's Health Initiative Investigators. (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women. J Am Med Assoc 288:321–333